B8

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁷
C12N 15/12
C12N 15/63 C07K 14/435
A61K 39/395 C12Q 1/68

[12]发明专利申请公开说明书

[21]申请号 99119883.2

[43]公开日 2001年7月11日

[11]公开号 CN 1302876A

[22]申请日 1999.10.28 [21]申请号 99119883.2

[71]申请人 复旦大学
地址 200433 上海市邯郸路220号
[72]发明人 余 龙 赵 勇 张宏来
傅 强 赵寿元

[74]专利代理机构 上海专利商标事务所
代理人 徐 迅

权利要求书1页 说明书24页 附图页数1页

[54]发明名称 人夏科－莱登晶体4及其编码序列，以及制法和用途

[57]摘要

本发明提供了人夏科－莱登晶体4(Charcot－Leyden Crystal4，简称为“CLC 4”)的cDNA序列，该cDNA编码的蛋白是人夏科－莱登晶体蛋白的同系物。本发明还涉及由该核苷酸序列编码的多肽，这些多核苷酸和多肽的应用，以及所述多核苷酸和所述多肽的生产方法。

ISSN1008-4274

知识产权出版社出版

权 利 要 求 书

1. 一种分离出的DNA分子，其特征在于，它包括：编码具有人CLC4蛋白活性的多肽的核苷酸序列，

所述的核苷酸序列与SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列有至少70%的同源性；或者

所述的核苷酸序列能在中度严紧条件下与SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列杂交.

2.如权利要求1所述的DNA分子，其特征在于，所述的序列编码一多肽，该多肽具有SEQ ID NO:4或12所示的序列.

3.如权利要求1所述的DNA分子，其特征在于，该序列具有SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列.

4.一种分离的人CLC4蛋白多肽，其特征在于，它包括：具有SEQ ID NO:4或12氨基酸序列的多肽、或其保守性变异多肽、或其活性片段、或其活性衍生物.

5.如权利要求4所述的多肽，其特征在于，该多肽是具有SEQ ID NO:4或12序列的多肽.

6.一种载体，其特征在于，它含有权利要求1所述的DNA.

7.一种用权利要求6所述载体转化的宿主细胞.

8.一种产生具有人CLC4蛋白活性的多肽的方法，其特征在于，该方法包括：

(a)将编码具有人CLC4蛋白活性的多肽的核苷酸序列可操作地连于表达调控序列，形成人CLC4蛋白表达载体，所述的核苷酸序列与SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列有至少70%的同源性；

(b)将步骤(a)中的表达载体转入宿主细胞，形成人CLC4蛋白的重组细胞；

(c)在适合表达人CLC4蛋白多肽的条件下，培养步骤(b)中的重组细胞；

(d)分离出具有人CLC4蛋白活性的多肽.

9.一种能与权利要求4所述的人CLC4蛋白多肽特异性结合的抗体.

10.一种探针分子，其特征在于，它含有权利要求 1 所述的 DNA 分子中 8-100 个连续核苷酸.

说　明　书

人夏科-莱登晶体4及其编码序列，以及制法和用途

本发明涉及基因工程领域，具体地，本发明涉及一种新的人基因核苷酸序列．更具体地说，本发明涉及人夏科-莱登晶体4(Charcot- Leyden Crystal 4, CLC4)的cDNA序列，该cDNA编码的蛋白是人夏科-莱登晶体蛋白的同系物．本发明还涉及由该核苷酸序列编码的多肽，这些多核苷酸和多肽的应用，以及所述多核苷酸和所述多肽的生产方法．

夏科-莱登晶体(Charcot- Leyden Crystal，CLC)是自然存在于人体组织和分泌物中的底面为六角的双锥体形晶体，与寄生过程、过敏性过程中外周血或组织中的嗜酸细胞数量增加有关．早在1854年，Charcot和Robin 首先在尸体的血液和白血病患者的脾脏中发现夏科-莱登晶体(Charcot, J. M.; Robin, C. : C.R. Seances Mem. Soc. Biol. Paris 5:44-52, 1854)，后来在一位哮喘病患者的唾液中也亦有发现(Leyden, E. Virchows Arch . Path. Anat. 54:324-344, 1872)．

CLC在嗜酸细胞相关炎症反应的组织和分泌物中被观察到，这些炎症反应有哮喘、骨髓性白血病、过敏反应、寄生虫病等(Beeson, P. B. 1977 et al. The eosinopil In Problems in the Internal Medicine, Vol. 15.; Ottesen, E, A. et al 1978. The eosinophil, eosinophilia and eosinophil-related disorders. In Allergy: Principles and Practice , Vol. 2. E. Middleton, C. E. Reed, and E. F. Ellis, eds. C. V. Mosby Co., St. Louis, MO, p. 584.)．CLC蛋白占嗜酸细胞总蛋白的约10%(Weller, P. F. et al. 1984, J. Biol. Chem. 259: 15100)，是嗜酸细胞的标志性蛋白，在体液和分泌物中鉴定到这种独特的底面为六角的双锥体形晶体—CLC晶体，可作为嗜酸细胞相关的过敏性炎症的标志．

CLC的全长cDNA序列在1993年分离得到，该cDNA序列长598bp，编码142氨基酸的多肽(Ackeman, S. J. et al 1993, J. Immun. 150: 456-468)．CLC的基因组序列几乎同时被获得，Dyer报道从特异的19号染色体库中克隆到包含了CLC编码序列的基因组片段，对CLC的cDNA和基因组序列分析结果显示CLC的编码顺序由4个外显子拼接而成(Dyer, K. D. et al. 1997,Genomics 40:17-221)．

Mastrianni将人夏科-莱登晶体蛋白定位于19号染色体(Mastrianni, D. M. et al 1991, Cytogenet. Cell Genet. 58 2021)．荧光原位杂交技术将CLC进一步定位于

19q13.1(B. Trask and H. Morhrenweiser unpublished observation, cited in Ackeman, S. J. et al 1993, J. Immun. 150: 456-468).

另有一些实验显示，夏科-莱登晶体作为一个独立的蛋白具有溶血磷脂酶的活性 (Gleich, G.J. et al 1976, J. Clin. Invest. 57: 633-640; Weller, P. F. et al 1980, Proc. Nat. Acad. Sci. U. S.A. 77; 7441-7443). 这一结论同时也被其它一些实验所支持: 碱性聚丙烯酰胺胶电泳显示，溶解的夏科-莱登晶体只有一条条带，且其分子量为17400，与从嗜酸性颗粒细胞中分离到的溶血磷脂酶的分子量相符；另外，嗜酸性颗粒细胞溶血磷脂酶能形成双锥型的晶体，与夏科-莱登晶体形态相同．这些都表明，人嗜酸粒细胞溶血磷脂酶是夏科-莱登晶体的组成成分(Weller, P. F. et al 1980, Proc. Nat. Acad. Sci. U. S.A. 77; 7441-7443). 但CLC与其它哺乳动物的溶血磷脂酶没有结构上的相似性(Leonidas, D. D. et al. 1995, Structure 3: 1379-1393).

而CLC蛋白与外源凝集素家族的某些成员在结构、部分功能上有相似性．CLC的cDNA推导蛋白(Ackeman, S. J. et al 1993, J. Immun. 150: 456-468)、基因组顺序的内含子-外显子的结构(Dyer, K. D. et al. 1997,Genomics 40:17-221)、特征性高度保守残基(Dyer, K. D. et al. 1996, Life Sci. 58: 2073-2082)、结合糖的活性(Dyer, K. D. et al. 1996, Life Sci. 58: 2073-2082)、晶体蛋白的整体结构(Leonidas, D. D. et al. 1995, Structure 3: 1379-1393)都与galectin家族的某些成员相似，这表明CLC蛋白可能具有galectin相关的一些活性，在细胞-细胞，细胞-基质的相互作用中通过β-半乳糖苷结合活性发挥生物功能(Dyer, K. D. et al. 1997, Gemonics 40: 217-221).

CLC是嗜酸细胞的标志，但在嗜碱粒细胞中也有CLC．亚显微结构研究表明，CLC存在于嗜酸性细胞和嗜碱粒细胞的颗粒中(Dvorak, A. M. et al. 1988, Blood 72: 150; Dvorak, A, M. et al. 1989, Lab. Invest. 60:557). 最近，通过免疫金技术或间接免疫荧光定位，还发现CLC存在于嗜酸性细胞和嗜碱粒细胞中的其他部位，如胞质和核(Dvorak, A. M. et al. 1990, Lab. Invest. 62:590-607; Dvorak, A. M. r tal. 1991, Am. J. Pathol. 138: 69-82; Dvorak, A. M. et al. 1997, Clin. Exp. Allergy 27: 452-474)、嗜碱粒细胞的脱粒通道、脱粒通道膜和一类与嗜酸性细胞的初级颗粒相似的颗粒．CLC蛋白分布的变化表明嗜碱粒细胞内含物(包括CLC)有释放与获得的变化过程，CLC的功能可能与其分布的变化过程相关(Dvorak, A. M. et al. 1997, Clin. Exp. Allergy 27: 452-474).

本发明的一个目的是提供一种新的多核苷酸，该多核苷酸被命名为人CLC4.

本发明的另一个目的是提供一种新的人CLC4蛋白.

本发明的再一个目的是提供一种利用重组技术生产所述的新的人的CLC4多肽的方法.

本发明还提供了这种人的CLC4核酸序列和多肽的应用.

在本发明的一个方面，提供了一种分离出的DNA分子，它包括：编码具有人CLC4蛋白活性的多肽的核苷酸序列，所述的核苷酸序列与SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列有至少70%的同源性；或者所述的核苷酸序列能在中度严紧条件下与SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列杂交．较佳地，所述的序列编码一多肽，该多肽具有SEQ ID NO:4或12所示的序列．更佳地，该序列具有SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列.

在本发明的另一方面，提供了一种分离的人CLC4蛋白多肽，它包括：具有SEQ ID NO:4或12氨基酸序列的多肽、或其保守性变异多肽、或其活性片段、或其活性衍生物．较佳地，该多肽是具有SEQ ID NO:4或12序列的多肽.

在本发明的另一方面，提供了一种载体，它含有上述分离出的DNA.

在本发明的另一方面，提供了一种所述载体转化的宿主细胞.

在本发明的另一方面，提供了一种产生具有人CLC4蛋白活性的多肽的方法，该方法包括：

(a)将编码具有人CLC4蛋白活性的多肽的核苷酸序列可操作地连于表达调控序列，形成人CLC4蛋白表达载体，所述的核苷酸序列与SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列有至少70%的同源性；

(b)将步骤(a)中的表达载体转入宿主细胞，形成人CLC4蛋白的重组细胞；

(c)在适合表达人CLC4蛋白多肽的条件下，培养步骤(b)中的重组细胞；

(d)分离出具有人CLC4蛋白活性的多肽.

在本发明的一个具体实施方案中，本发明的分离的多核苷酸全长为452个核苷酸，其详细序列见SEQ ID NO:3，其中开放读框位于11-427位核苷酸它编码一种人CLC4蛋白-即CLC4A蛋白．在本发明的另一具体实施方案中，本发明的分离的多核苷酸全长为662个核苷酸，其详细序列见SEQ ID NO. 11，其中开放读框位于95-598位核苷酸，它编码另一种人CLC4蛋白-即CLC4B蛋白．CLC4A和CLC4B是同一基

因经不同剪接后所产生的蛋白.

在本发明中，“分离的”、“纯化的”或“基本纯的”DNA是指，该DNA或片段已从天然状态下位于其两侧的序列中分离出来，还指该DNA或片段已经与天然状态下伴随核酸的组份分开，而且已经与在细胞中伴随其的蛋白质分开.

在本发明中，术语“人CLC4蛋白(或多肽)编码序列”指编码具有人CLC4蛋白活性的多肽的核苷酸序列，如SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列及其简并序列. 该简并序列是指，位于SEQ ID NO:3序列的编码框11-427位核苷酸中或位于SEQ ID NO:11序列的编码框95-598位核苷酸中，有一个或多个密码子被编码相同氨基酸的简并密码子所取代后而产生的序列. 由于密码子的简并性，所以与SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列同源性低至约70%的简并序列也能编码出SEQ ID NO:4或12所述的序列. 该术语还包括能在中度严紧条件下，更佳地在高度严紧条件下，与SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列杂交的核苷酸序列. 还术语还包括与SEQ ID NO. 3中从核苷酸11-427位的核苷酸序列或SEQ ID NO. 11中从核苷酸95-598位的核苷酸序列的同源性至少70%，较佳地至少80%，更佳地至少90%的核苷酸序列.

该术语还包括能编码具有与人CLC4相同功能的蛋白的、SEQ ID NO:4或12序列的变异形式. 这些变异形式包括(但并不限于): 若干个(通常为1-90个，较佳地1-60个，更佳地1-20个，最佳地1-10个)核苷酸的缺失、插入和/或取代，以及在5 和/或3 端添加数个核苷酸. 本发明的编码序列可以是DNA或RNA，可以是单链或双链.

在本发明中，“基本纯的”蛋白质或多肽是指其至少占样品总物质的至少20%，较佳地至少50%，更佳地至少80%，最佳地至少90%(按干重或湿重计). 纯度可以用任何合适的方法进行测量，如用柱层析、PAGE或HPLC法测量多肽的纯度. 基本纯的多肽基本上不含天然状态下的伴随其的组分.

在本发明中，术语“人CLC4蛋白多肽”指具有天然人CLC4蛋白活性的SEQ ID NO:4或12序列的多肽. 该术语还包括具有与人CLC4相同功能的、SEQ ID NO:4或12序列的变异形式. 这些变异形式包括(但并不限于): 若干个(通常为1-50个，较佳地1-30个，更佳地1-20个，最佳地1-10个)氨基酸的缺失、插入和/或取代，以及在C末端和/或N末端添加一个或数个(通常为20个以内，较佳地为10个以内，更佳地为5个以内)氨基酸. 例如，在本领域中，用性能相近或相似的氨基酸进行取代时，通常不会改变蛋白质的功能. 又比如，在C末端和/或N末端添加一个或数个氨基酸通

常也不会改变蛋白质的功能. 该术语还包括人CLC4蛋白的活性片段和活性衍生物.

该多肽的变异形式包括: 同源序列、保守性变异体、等位变异体、天然突变体、诱导突变体、在高或低的严紧度条件下能与人CLC4 DNA 杂交的DNA所编码的蛋白、以及利用抗人CLC4多肽的抗血清获得的多肽或蛋白. 本发明还提供了其他多肽, 如包含人CLC4多肽或其片段的融合蛋白. 除了几乎全长的多肽外, 本发明还包括了人CLC4多肽的可溶性片段. 通常, 该片段具有人CLC4多肽序列的至少约10个连续氨基酸, 通常至少约30个连续氨基酸, 较佳地至少约50个连续氨基酸, 更佳地至少约80个连续氨基酸, 最佳地至少约100个连续氨基酸.

发明还提供人CLC4蛋白或多肽的类似物. 这些类似物与天然人CLC4多肽的差别可以是氨基酸序列上的差异, 也可以是不影响序列的修饰形式上的差异, 或者兼而有之. 这些多肽包括天然或诱导的遗传变异体. 诱导变异体可以通过各种技术得到, 如通过辐射或暴露于诱变剂而产生随机诱变, 还可通过定点诱变法或其他已知分子生物学的技术. 类似物还包括具有不同于天然L-氨基酸的残基(如D-氨基酸)的类似物, 以及具有非天然存在的或合成的氨基酸(如β、γ-氨基酸)的类似物. 应理解, 本发明的多肽并不限于上述例举的代表性的多肽.

修饰(通常不改变一级结构)形式包括: 体内或体外的多肽的化学衍生形式如乙酰化或羧基化. 修饰还包括糖基化, 如那些在多肽的合成和加工中或进一步加工步骤中进行糖基化修饰而产生的多肽. 这种修饰可以通过将多肽暴露于进行糖基化的酶(如哺乳动物的糖基化酶或去糖基化酶)而完成. 修饰形式还包括具有磷酸化氨基酸残基(如磷酸酪氨酸, 磷酸丝氨酸, 磷酸苏氨酸)的序列. 还包括被修饰从而提高了其抗蛋白水解性能或优化了溶解性能的多肽.

在本发明中, "人CLC4保守性变异多肽" 指与SEQ ID NO. 4或12的氨基酸序列相比, 有至多10个, 较佳地至多8个, 更佳地至多5个, 最佳地至多3个氨基酸被性质相似或相近的氨基酸所替换而形成多肽. 这些保守性变异多肽最好根据表1进行氨基酸替换而产生.

表 1

最初的残基	代表性的取代	优选的取代
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Lys; Arg	Gln

Asp (D)	Glu	Glu
Cys (C)	Ser	Ser
Gln (Q)	Asn	Asn
Glu (E)	Asp	Asp
Gly (G)	Pro; Ala	Ala
His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe	Leu
Leu (L)	Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Leu; Val; Ile; Ala; Tyr	Leu
Pro (P)	Ala	Ala
Ser (S)	Thr	Thr
Thr (T)	Ser	Ser
Trp (W)	Tyr; Phe	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala	Leu

本发明还包括人CLC4多肽编码序列及其片段的反义序列．这种反义序列可用于抑制细胞内人CLC4的表达．

本发明还包括一种可用作探针的核酸分子，该分子通常具有人CLC4多肽编码序列的8-100个，较佳地15-50个连续核苷酸．该探针可用于检测样品中是否存在编码人CLC4的核酸分子．

本发明还包括检测人CLC4核苷酸序列的方法，它包括用上述的探针与样品进行杂交，然后检测探针是否发生了结合．较佳地，该样品是PCR扩增后的产物，其中PCR扩增引物对应于人CLC4多肽的编码序列，并可位于该编码序列的两侧或中间．引物长度一般为15-50个核苷酸．

在本发明中，可选用本领域已知的各种载体，如市售的载体．比如，选用市售的载体，然后将编码本发明多肽的核苷酸序列可操作地连于表达调控序列，可以形成蛋白表达载体．

如本文所用，“可操作地连于”指这样一种状况，即线性DNA序列的某些部分能够影响同一线性DNA序列其他部分的活性．例如，如果信号肽DNA作为前体表达并参与多肽的分泌，那么信号肽(分泌前导序列)DNA就是可操作地连于多肽DNA；如果启动子控制序列的转录，那么它是可操作地连于编码序列；如果核糖体结合位点被置于能使其翻译的位置时，那么它是可操作地连于编码序列．一般，“可操作地连于” 意味着相邻近，而对于分泌前导序列则意味着在阅读框中相邻．

在本发明中，术语“宿主细胞”包括原核细胞和真核细胞．常用的原核宿主细胞的例子包括大肠杆菌、枯草杆菌等．常用的真核宿主细胞包括酵母细胞，昆虫细胞、和哺乳动物细胞．较佳地，该宿主细胞是真核细胞，如CHO细胞、COS细胞等．

另一方面，本发明还包括对人CLC4 DNA或是其片段编码的多肽具有特异性的多克隆抗体和单克隆抗体，尤其是单克隆抗体．这里，“特异性”是指抗体能结合于人CLC4基因产物或片段．较佳地，指那些能与人CLC4基因产物或片段结合但不识别和结合于其它非相关抗原分子的抗体．本发明中抗体包括那些能够结合并抑制人CLC4蛋白的分子，也包括那些并不影响人CLC4蛋白功能的抗体．本发明还包括那些能与修饰或未经修饰形式的人CLC4基因产物结合的抗体．

本发明不仅包括完整的单克隆或多克隆抗体，而且还包括具有免疫活性的抗体片段，如Fab'或$(Fab)_2$片段；抗体重链；抗体轻链；遗传工程改造的单链Fv分子(Ladner等人，美国专利No. 4,946,778)；或嵌合抗体，如具有鼠抗体结合特异性但仍保留来自人的抗体部分的抗体．

本发明的抗体可以通过本领域内技术人员已知的各种技术进行制备．例如，纯化的人CLC4基因产物或者其具有抗原性的片段，可被施用于动物以诱导多克隆抗体的产生．与之相似的，表达人CLC4或其具有抗原性的片段的细胞可用来免疫动物来生产抗体．本发明的抗体也可以是单克隆抗体．此类单克隆抗体可以利用杂交瘤技术来制备(见 Kohler等人, Nature 256;495, 1975; Kohler等人, Eur.J.Immunol. 6:511, 1976; Kohler 等人, Eur.J.Immunol. 6:292, 1976; Hammerling 等人, In Monoclonal Antibodies and T Cell Hybridomas, Elsevier, N.Y., 1981)．本发明的抗体包括能阻断人CLC4功能的抗体以及不影响人CLC4功能的抗体．本发明的各类抗体可以利用人CLC4基因产物的片段或功能区，通过常规免疫技术获得．这些片段或功能区可以利用重组方法制备或利用多肽合成仪合成．与人CLC4基因产物的未修饰形式结合的抗体可以用原核细胞(例如*E. Coli*)中生产的基因产物来免疫动物而产

生；与翻译后修饰形式结合的抗体(如糖基化或磷酸化的蛋白或多肽)，可以用真核细胞(例如酵母或昆虫细胞)中产生的基因产物来免疫动物而获得.

本发明的人CLC4核苷酸全长序列或其片段通常可以用PCR扩增法、重组法或人工合成的方法获得. 对于PCR扩增法，可根据本发明所公开的有关核苷酸序列，尤其是开放阅读框序列来设计引物，并用市售的cDNA库或按本领域技术人员已知的常规方法所制备的cDNA库作为模板，扩增而得有关序列. 当序列较长时，常常需要进行两次或多次PCR扩增，然后再将各次扩增出的片段按正确次序拼接在一起.

一旦获得了有关的序列，就可以用重组法来大批量地获得有关序列. 这通常是将其克隆入载体，再转入细胞，然后通过常规方法从增殖后的宿主细胞中分离得到有关序列.

此外，还可用人工合成的方法来合成有关序列，尤其是片段长度较短时. 通常，通过先合成多个小片段，然后再进行连接可获得序列很长的片段.

目前，已经可以完全通过化学合成来编码本发明蛋白(或其片段，或其衍生物)的 DNA 序列. 然后可将该 DNA 序列引入本领域中的各种 DNA 分子(如载体)和细胞中. 此外，还可通过化学合成将突变引入本发明蛋白序列中.

本发明蛋白的片段除了可用重组法产生之外，还可用固相技术通过直接合成肽而加以生产(Stewart 等人,(1969) Solid-Phase Peptide Synthesis, WH Freeman Co., San Francisco; Merrifield J. (1963) J. Am Chem. Soc 85: 2149-2154). 在体外合成蛋白质可以用手工或自动进行. 例如，可以用 Applied Biosystems 的 431A 型肽合成仪(Foster City, CA)来自动合成肽. 可以分别化学合成本发明蛋白的各片段，然后用化学方法加以连接以产生全长的分子.

本发明蛋白的编码序列还可用于基因定位. 例如，通过荧光原位杂交技术(FISH)，将 cDNA 克隆与分裂中期的染色体进行杂交，可以准确地进行染色体定位. 该技术可以使用短至约 500bp 的 cDNA；也可以使用长至约 2000bp 或者更长的 cDNA . 对于该技术，可参见 Verma 等人, Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York(1988).

一旦序列被定位于染色体上的某个精确位置，将可以将序列在染色体上的物理位置与遗传图谱数据相关联. 这些遗传图谱数据是可以获得的，例如通过孟德尔(Mendelian)人遗传数据库(可通过 Johns Hopkins University Welch Medical Library 在网上获得). 然后，通过连锁分析来鉴定基因与已定位于同一染色体区域的疾病之间的相关性.

接着，有必要确定患病个体和健康个体之间的 cDNA 或基因组序列方面的差异．如果某一突变存在于部分或全部患病个体但不存在于正常个体，那么该突变可能就是该疾病的致病因素．

此外，利用本发明蛋白，通过各种常规筛选方法，可筛选出与 CLC4 发生相互作用的物质，如受体、抑制剂或拮抗剂等．

本发明蛋白及其抗体、抑制剂、拮抗剂或受体等，当在治疗上进行施用(给药)时，可提供不同的效果．通常，可将这些物质配制于无毒的、惰性的和药学上可接受的水性载体介质中， 其中 pH 通常约为 5-8，较佳地 pH 约为 6-8，尽管 pH 值可随被配制物质的性质以及待治疗的病症而有所变化．配制好的药物组合物可以通过常规途径进行给药，其中包括(但并不限于)：肌内、腹膜内、皮下、皮内、或局部给药．

以本发明的人 CLC4 蛋白为例，可以将其与合适的药学上可接受的载体联用．这类药物组合物含有治疗有效量的蛋白质和药学上可接受的载体或赋形剂．这类载体包括(但并不限于)：盐水、缓冲液、葡萄糖、水、甘油、乙醇、及其组合．药物制剂应与给药方式相匹配．本发明的人 CLC4 蛋白可以被制成针剂形式，例如用生理盐水或含有葡萄糖和其他辅剂的水溶液通过常规方法进行制备．诸如片剂和胶囊之类的药物组合物，可通过常规方法进行制备．药物组合物如针剂、溶液、片剂和胶囊宜在无菌条件下制造．活性成分的给药量是治疗有效量，例如每天约 1 微克/千克体重-约 5 毫克/千克体重．此外，本发明的多肽还可与其他治疗剂一起使用．

当本发明的人CLC4蛋白多肽被用作药物时，可将治疗有效剂量的该多肽施用于哺乳动物，其中该治疗有效剂量通常至少约10微克/千克体重，而且在大多数情况下不超过约8毫克/千克体重，较佳地该剂量是约10微克/千克体重-约1毫克/千克体重．当然，具体剂量还应考虑给药途径、病人健康状况等因素，这些都是熟练医师技能范围之内的．

CLC与嗜酸细胞相关炎症反应有着密切的关系，其在体液和分泌物中的鉴定可作为嗜酸细胞相关的过敏性炎症的标志．另外，CLC与外源凝集素家族的某些成员在结构、部分功能的相似性表明CLC蛋白可能具有外源凝集素相关的一些活性，在细胞-细胞，细胞-基质的相互作用中通过β-半乳糖苷结合活性发挥生物功能(Dyer, K. D. et al. 1997, Gemonics 40: 217-221)．CLC还存在于嗜碱粒细胞中，其分布有变化的过程，CLC的功能可能与其分布的变化过程相关(Dvorak, A. M. et al. 1997, Clin. Exp. Allergy 27: 452-474).

此外，由于本发明的人 CLC4 具有源自人的天然氨基酸序列，因此，与来源于其他物种的同族蛋白相比，预计在施用于人时将具有更高的活性和/或更低的副作用(例如在人体内的免疫原性更低或没有).

在附图中，图1为本发明的人CLC4A、CLC4B与人CLC(hCLC)、鼠CLC(mCLC)和猩猩CLC(pCLC)等5种蛋白的氨基酸序列同源比较图．其中，相同的氨基酸在序列下方用“*”标出，相似的氨基酸用“•”标出.

下面结合具体实施例，进一步阐述本发明．应理解，这些实施例仅用于说明本发明而不用于限制本发明的范围．下列实施例中未注明具体条件的实验方法，通常按照常规条件如Sambrook等人， 分子克隆: 实验室手册(New York: Cold Spring Harbor Laboratory Press, 1989)中所述的条件，或按照制造厂商所建议的条件.

实施例1

人CLC4的cDNA序列的克隆和测定

(A)人CLC4A的cDNA序列的克隆和测定

1. 引物扩增

以人脑λ gt11cDNA文库(购自Clontech公司)为模板，用一对寡核苷酸为引物——A: 5'-AAGAGAGACAATGTCTCTTAACC-3' (SEQ ID NO:1)为正向引物，寡核苷酸B: 5'- ACAATGAGGAGTGTGATCATCTC -3' (SEQ ID NO:2)为反向引物，进行PCR．PCR条件为93℃4分钟，随之以93℃1分钟、66℃1分钟和72℃1分钟进行35个循环，最后72℃延伸5分钟．电泳检测得到的PCR片断约为450bp的目的片段.

2. PCR产物的测序

将上述PCR扩增产物AB与pGEM-T 载体(Promega)连接，转化大肠杆菌JM103，用QIAprep Plasmid试剂盒(QIAGEN)提取质粒，用SequiTherm EXCEL™ DNA测序试剂盒(Epicentre Technologies)对插入片段进行测序，最后用电脑软件拼接顺序，获得全长cDNA序列，共452bp，详细序列见SEQ ID NO:3，其中开放读框位于11-427位核苷酸.

根据得到的全长cDNA序列推导出人CLC4的氨基酸序列，共138个氨基酸残基，其氨基酸序列详见SEQ ID NO:4.

(B)人CLC4B的cDNA序列的克隆和测定

1. 引物扩增

采用与克隆CLC4A相同的方法，不同点仅在于：所用的一对寡核苷酸引物不同，即将寡核苷酸A'：5'- GCCTGGGTGACAGAGCAAGAACC -3' (SEQ ID NO:9)为正向引物，寡核苷酸B'：5'-CTGGGAATCCCATGGTCAGGTAG-3' (SEQ ID NO:10)为反向引物，进行PCR．电泳检测得到的PCR片段约为650bp的目的片段．

2. PCR产物的测序

如上将上述PCR扩增产物A'/B'与pGEM-T®载体(Promega)连接，转化大肠杆菌JM103，用QIAprep Plasmid试剂盒(QIAGEN)提取质粒，用SequiTherm EXCEL™DNA测序试剂盒(Epicentre Technologies)对插入片段进行测序，获得全长cDNA序列，共662bp，详细序列见SEQ ID NO: 11，其中开放读框位于95-598位核苷酸．

根据得到的全长cDNA序列推导出人CLC4B的氨基酸序列，共167个氨基酸残基，其氨基酸序列详见SEQ ID NO:12．

实施例2

同源比较

用本发明的人CLC4的全长cDNA序列及其编码蛋白，在Non-redundant GenBank+ EMBL+DDBJ+PDB 数据库及Non-redundant GenBank CDS translations+PDB+ SwissProt+Spupdate+PIR数据库中，用BLAST程序进行核酸和蛋白同源检索．结果发现，在核苷酸水平和蛋白水平上与人的CLC(gb|L01664)都有较高同源性．其中，CLC4A推导蛋白与人CLC 蛋白的同一性达到54%，相似性达61%．CLC4B推导蛋白与人CLC 蛋白的同一性达到53%，相似性达59%．另外，CLC4A和CLC4B与鼠和猩猩中已公布的CLC部分序列也显示了较高的同源性(CLC4A和CLC4B与hCLC、mCLC及pCLC的氨基酸序列同源比较见图1)．

此外，CLC4A和CLC4B蛋白几乎相同，不同点仅在于：CLC4B在N端稍不同，即多出了近30个氨基酸(图1)．因此，这暗示CLC4A和CLC4B是同一基因的不同剪接本．

根据结构决定功能，结构相似功能相关的生物学原理，可以根据已知的不同物种中的CLC基因及其编码蛋白的功能来推测本发明CLC4的生物学功能．

CLC在嗜酸细胞相关炎症反应的组织和分泌物中观察到，这些炎症反应有哮喘，骨髓白血病，过敏反应，寄生虫病等(Beeson, P. B. 1977 et al. The eosinopil In Problems in the Internal Medicine, Vol. 15.; Ottesen, E, A. et al 1978. The eosinophil, eosinophilia and eosinophil-related disorders. In Allergy: Principles and Practice , Vol.

2. E. Middleton, C. E. Reed, and E. F. Ellis, eds. C. V. Mosby Co., St. Louis, MO, p. 584.). CLC蛋白占嗜酸细胞总蛋白的约10%(Weller, P. F. et al. 1984, J. Biol. Chem. 259: 15100)，是嗜酸细胞的标志性蛋白，在体液和分泌物中鉴定到这种独特的底面为六角的双锥体形晶体—CLC晶体，可作为嗜酸细胞相关的过敏性炎症的标志.

而CLC蛋白与外源凝集素家族的某些成员存在结构、部分功能上的相似性. CLC的cDNA推导蛋白与结合β-半乳糖苷S-样动物外源凝集素(galectin)超家族的成员有相似性(Ackeman, S. J. et al 1993, J. Immun. 150: 456-468). CLC基因组顺序的内含子-外显子的结构与galectin的结构相似，全部β-半乳糖苷结合位点由一个外显子-编码，这一特征与目前为止所有的galectin相符合(Dyer, K. D. et al. 1997,Genomics 40:17-221).

CLC蛋白和人galectin的蛋白序列分析表明，CLC4A与该家族的特征性的13个高度保守残基中的11个相同或相似，它们是第23位的G、第52位的R、第54位的R、第56位的H、第62位的V、第64位的N、第71位的W、第74位的E、第82位的F、第85位的G和第114位的R. CLC4B与该家族的特征性的13个高度保守残基中的11个相同或相似，它们是第52位的G、第81位的R、第83位的R、第85位的H、第97位的V、第93位的N、第100位的W、第103位的E、第111位的F、第114位的G和第143位的R.

CLC蛋白能结合于偶合乳糖的琼脂糖树脂，而且这种结合作用还依赖于糖的剂量(Dyer, K. D. et al. 1996, Life Sci. 58: 2073-2082). X-射线晶体衍射表明，CLC晶体蛋白的整体结构与galectin-1和-2高度相似. CLC含有保守的galectin的碳水化合物识别结构域的大部分残基(Leonidas, D. D. et al. 1995, Structure 3: 1379-1393). 这些结果都表明CLC蛋白可能具有外源凝集素相关的一些活性，可能在细胞-细胞，细胞-基质的相互作用中通过β-半乳糖苷结合活性发挥生物功能(Dyer, K. D. et al. 1997, Gemonics 40: 217-221).

CLC是嗜酸细胞的标志，但在嗜碱粒细胞中也有CLC. 亚显微结构研究表明，CLC存在于嗜酸性细胞和嗜碱粒细胞的颗粒中(Dvorak, A. M. et al. 1988, Blood 72: 150; Dvorak, A, M. et al. 1989, Lab. Invest. 60:557). 最近，通过免疫金技术或间接免疫荧光定位，还发现CLC存在于嗜酸性细胞和嗜碱粒细胞中的其他部位，如胞质和核(Dvorak, A. M. et al. 1990, Lab. Invest. 62:590-607; Dvorak, A. M. r tal. 1991, Am. J. Pathol. 138: 69-82; Dvorak, A. M. et al. 1997, Clin. Exp. Allergy 27: 452-474)、嗜碱粒细胞的脱粒通道、脱粒通道膜和一类与嗜酸性细胞的初级颗粒相似的颗粒. CLC蛋白分布的变化表明嗜碱粒细胞内含物(包括CLC)有释放与获得的过程变

化，CLC的功能可能与其分布的变化过程相关(Dvorak, A. M. et al. 1997, Clin. Exp. Allergy 27: 452-474).

本发明的人CLC4除了可作为该家族一员用于进一步的功能研究，还可用于与其他蛋白一起产生融合蛋白，比如与免疫球蛋白一起产生融合蛋白．此外，本发明人CLC4还可以与该家族的其他成员进行融合或交换片段，以产生新的蛋白．例如将本发明人CLC4的N端与鼠CLC的N端进行交换，以产生新的活性更高或具有新特性的蛋白．

本发明的人CLC4还可用于筛选抑制本发明蛋白活性的拮抗剂、或增强本发明蛋白活性的激动剂等．

针对本发明人CLC4的抗体，用于筛选该家族的其他成员，或者用于亲和纯化相关蛋白(如该家族的其他成员)．

此外，本发明人CLC4核酸(编码序列或反义序列)可以被引入细胞，以提高人CLC4的表达水平或者抑制人CLC4的过度表达．本发明的人CLC4蛋白或其活性多肽片段可以施用于病人，以治疗或减轻因人CLC4缺失、无功能或异常而导致的有关病症．此外，还可以用基于本发明的核酸序列或抗体进行有关的诊断或预后判断．

实施例3

(A)人CLC4A在大肠杆菌中的表达

在该实施例中，以实施例1 中PCR扩增产物A/B为模板，将编码人CLC4A的cDNA序列用对应于该DNA序列的5'和3'端的PCR寡核苷酸引物(SEQ ID NO:5和6)进行扩增，获得人CLC4A cDNA作为插人片段．

PCR反应中使用的5'寡核苷酸引物序列为:

5 - CAACGTCGACATGTCTCTTAACCGTGCAT -3'(SEQ ID NO:5)，该引物含有SalI限制性内切酶的酶切位点，在该酶切位点之后是由起始密码子开始的人CLC4A编码序列的19个核苷酸;

3'端引物序列为:

5 - TAGGAAGCTTTCAATTGCAGACACACACT-3 (SEQ ID NO:6)，该引物含有HindIII限制性内切酶的酶切位点、翻译终止子和人CLC4A的部分编码序列．

引物上的限制性内切酶的酶切位点对应于细菌表达载体pQE-9(Qiagen Inc., Chatsworth, CA)上的限制性内切酶酶切位点，该质粒载体编码抗生素抗性(Amp^r)、

一个细菌复制起点(ori)、一个IPTG-可调启动子/操纵子(P/O)、一个核糖体结合位点(RBS)、一个6-组氨酸标记物(6-His)以及限制性内切酶克隆位点.

用SalI和HindIII消化pQE-9载体及插入片段，随后将插入片段连接到pQE-9载体并保持开放读框在细菌RBS起始．随后用连接混合物转化购自Qiagen，商品名为M15/rep4的E.coli菌株，M15/rep4含有多拷贝的质粒pREP4，其表达lacI阻遏物并携带卡那霉素抗性(Kan^r)．在含有Amp和Kan的LB培养皿上筛选转化子，抽提质粒测序验证人CI.C4A的cDNA片段已正确插入了载体.

在补加Amp(100 μ g/ml)和Kan(25 μ g/ml)的LB液体培养基中过夜培养(O/N)含所需构建物的阳性转化子克隆．过夜(O/N)培养物以1：100-1：250的稀释率稀释，然后接种到大体积培养基中，培养细胞生长至600光密度(OD_{600})为0.4-0.6时，加入IPTG(“异丙基硫代-β-D-半乳糖苷”)至终浓度为1mM．通过使lacI阻遏物失活，IPTG诱导启动P/O导致基因表达水平提高．继续培养细胞3-4小时，随后离心(6000 × g，20分钟)．超声裂解培养物，收集细胞裂解液并将其稀释于6M盐酸胍中．澄清后，通过在能使含6-His标记物蛋白紧密结合的条件下，用镍-螯合柱层析从溶液中纯化溶解的人CLC4 ．用6M盐酸胍(pH5.0)从柱中洗脱人CLC4．可用几种方法从盐酸胍中变性沉淀蛋白．或者，使用透析步骤除去盐酸胍，或者从镍-螯合柱中分离出的纯化蛋白．纯化后的蛋白质被结合到第二个柱中，该柱中具有递减的线性盐酸胍梯度．在结合到该柱时蛋白质变性，随后用盐酸胍(pH5.0)洗脱．最后，将可溶的蛋白质用PBS进行透析,然后将蛋白质保存在终浓度为10%(w/v)甘油的贮存液中.

用12%的SDS-PAGE胶进行电泳，鉴定表达蛋白的分子量大小约为16kDa.

此外，用常规方法对表达蛋白CLC4A的N端和C端各10个氨基酸长度的氨基酸进行测序，发现与SEQ ID NO:4的序列一致.

(B)人CLC4B在大肠杆菌中的表达

采用与实施例3(A)相同的程序对人CLC4B进行表达，不同点仅在于：以实施例1中PCR扩增产物A'/B'为模板，用对应于人CLC4B的cDNA序列5'和3'端的PCR寡核苷酸引物(SEQ ID NO.13和14)代替SEQ ID NO: 5和6所示的引物.

PCR反应中使用的5'寡核苷酸引物序列为:

5 - CAACGGATCCATGTCCCTGACCCACAAGC -3'(SEQ ID NO:13),

该引物含有BamHI限制性内切酶的酶切位点，在该酶切位点之后是由起始密码子开始的人CLC4B编码序列的20个核苷酸;

3'端引物序列为:

5 - TAGGGTCGACTCAATTGCAGACACACACT-3 (SEQ ID NO:14), 该引物含有SalI限制性内切酶的酶切位点、翻译终止子和人CLC4B的部分编码序列.

然后，用SalI和HindIII消化pQE-9载体及插入片段，随后将插入片段连接到pQE-9载体并保持开放读框在细菌RBS起始. 随后用连接混合物转化购自Qiagen, 商品名为M15/rep4的E.coli菌株. 在含有Amp和Kan的LB培养皿上筛选转化子，抽提质粒，测序验证人CLC4B的cDNA片段已正确插入了载体.

获得的表达蛋白CLC4B用12%的SDS-PAGE胶进行电泳，鉴定表达蛋白的分子量大小约为19kDa. 用常规方法对人CLC4B蛋白的N端和C端各10个氨基酸长度的氨基酸进行测序，发现与SEQ ID NO. 12的序列一致.

实施例4

(A)人CLC4A在真核细胞(CHO细胞株)中的表达

在该实施例中，以实施例1 中PCR扩增产物A/B为模板，将编码人CLC4A的cDNA序列用对应于该DNA序列的5'和3'端的PCR寡核苷酸引物(SEQ ID NO:7和8)进行扩增，获得人CLC4A cDNA作为插入片段.

PCR反应中使用的5'寡核苷酸引物序列为:

5'- CAACAAGCTTATGTCTCTTAACCGTGCAT -3' (SEQ ID NO:7)

该引物含有HindIII限制性内切酶的酶切位点，在该酶切位点之后是由起始密码子开始的人CLC4A编码序列的19个核苷酸;

3'端引物序列为:

5 - TAGGGGATCCTCAATTGCAGACACACACT -3 (SEQ ID NO:8)

该引物含有BamHI限制性内切酶的酶切位点、一个翻译终止子和人CLC4A的部分编码序列.

引物上的限制性内切酶的酶切位点对应于CHO细胞表达载体pcDNA3上的限制性内切酶酶切位点，该质粒载体编码抗生素抗性(Amp^r和Neo^r)、一个噬菌体复制起点(f1 ori)、一个病毒复制起点(SV40 ori)、一个T7启动子、一个病毒启动子(P-CMV)、一个Sp6启动子、一个SV40启动子、一个SV40加尾信号和相应的polyA顺序、一个BGH加尾信号和相应的polyA顺序.

用HindIII和BamHI消化pcDNA3载体及插入片段，随后将插入片段连接到pcDNA3载体. 随后用连接混合物转化E.coli DH5α菌株. 在含有Amp的LB培养皿

上筛选转化子，在补加Amp(100μg/ml)的LB液体培养基中过夜培养(O/N)含所需构建物的克隆．抽提质粒测序验证人CLC4A的cDNA片段已正确插入了载体．

质粒转染CHO细胞是采用脂转染法，用Lipofectin (GiBco Life)进行的．转染48小时后，经2-3周的持续G418加压筛选，收集细胞及细胞上清测定表达蛋白酶活力．去G418，连续传代培养；对混合克隆细胞极限稀释，选择具有较高蛋白活性的细胞亚克隆．按常规方法大量培养上述阳性亚克隆．48小时后，开始收集细胞及上清，用超声裂解方法破碎细胞．以含0.05%Triton的50mM Tris.HCl(pH7.6)溶液为平衡液及洗脱液，用经预平衡的Superdex G-75柱收集上述蛋白的活性峰．再用50mMTris.HCl(pH8.0)平衡的DEAE-Sepharose柱，以含0-1M NaCl的50mMTris.HCl(pH8.0)溶液为洗脱液进行梯度洗脱，收集上述蛋白的活性峰．然后以PBS(pH7.4)为透析液对表达蛋白溶液进行透析．最后冻干保存．

用12%的SDS-PAGE胶进行电泳，鉴定表达蛋白的分子量大小为16kDa．

此外，用常规方法对表达蛋白CLC4A的N端和C端各10个氨基酸长度的氨基酸进行测序，发现与SEQ ID NO:4的序列一致．

(B)人CLC4B在真核细胞(CHO细胞株)中的表达

采用与实施例4(A)相同的程序对人CLC4B进行表达，不同点仅在于：以实施例1中PCR扩增产物A'/B'为模板，用对应于人CLC4B的cDNA序列5'和3'端的PCR寡核苷酸引物(SEQ ID NO.15和16)代替SEQ ID NO: 7和8所示的引物．

PCR反应中使用的5'寡核苷酸引物序列为:

5'- CAACGGATCCATGTCCCTGACCCACAAGC -3' (SEQ ID NO:15)

该引物含有BamHI限制性内切酶的酶切位点，在该酶切位点之后是由起始密码子开始的人CLC4B编码序列的19个核苷酸;

3'端引物序列为:

5 - TAGGGAATTCTCAATTGCAGACACACACT -3 (SEQ ID NO:16)

该引物含有EcoRI限制性内切酶的酶切位点、一个翻译终止子和人CLC4B的部分编码序列．

然后，用HindIII和BamHI消化pcDNA3载体及插入片段，随后将插入片段连接到pcDNA3载体．随后用连接混合物转化E.coli DH5α菌株．如上所述．在含有Amp的LB培养皿上筛选转化子，在补加Amp(100μg/ml)的LB液体培养基中过夜培养(O/N)含所需构建物的克隆．抽提质粒，测序验证人CLC4B的cDNA片段已正确插入了载体．

如上用质粒转染CHO细胞、表达蛋白并分离纯化．获得的蛋白用12%的SDS-PAGE胶进行电泳，鉴定表达蛋白的分子量大小为19kDa．用常规方法对人CLC4B蛋白的N端和C端各10个氨基酸长度的氨基酸进行测序，发现与SEQ ID NO. 12的序列一致．

实施例5

制备抗体

将实施例3和4中获得的CLC4A或CLC4B重组蛋白，分别用来免疫动物以产生抗体，具体方法如下．重组分子用层析法进行分离后备用．也可用SDS-PAGE凝胶电泳法进行分离，将电泳条带从凝胶中切下，并用等体积的完全Freund's佐剂乳化．用50-100μg/0.2ml乳化过的蛋白，对小鼠进行腹膜内注射．14天后，用非完全Freund's佐剂乳化的同样抗原，对小鼠以50-100μg/0.2ml的剂量进行腹膜内注射以加强免疫．每隔14天进行一次加强免疫，至少进行三次．获得的抗血清的特异反应活性用它在体外沉淀人CLC4基因翻译产物的能力加以评估．结果发现，抗体可特异性地与本发明蛋白发生沉淀．

在本发明提及的所有文献都在本申请中引用作为参考，就如同每一篇文献被单独引用作为参考那样．此外应理解，在阅读了本发明的上述讲授内容之后，本领域技术人员可以对本发明作各种改动或修改，这些等价形式同样落于本申请所附权利要求书所限定的范围．

序 列 表

(1)一般信息:

(i)申请人: 复旦大学

(ii)发明名称: 人夏科-莱登晶体4及其编码序列，以及制法和用途

(iii)序列数目: 16

(2)SEQ ID NO:1的信息

(i)序列特征

(A)长度: 23碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:1:

AAGAGAGACA ATGTCTCTTA ACC 23

(2)SEQ ID NO:2的信息

(i)序列特征

(A)长度: 23碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:2

ACAATGAGGA GTGTGATCAT CTC 23

(2)SEQ ID NO:3的信息:

(i)序列特征:

(A)长度: 452bp

(B)类型: 核酸

(C)链性: 双链

(D)拓扑结构: 线性

(ii)分子类型: cDNA

(xi)序列描述: SEQ ID NO:3

AAGAGAGACA ATGTCTCTTA ACCGTGCATA CAAACTGCCT GTGTCTTTGT CTGTTGGTTC 60
TTGCGTGATA ATCAAAGGGA CACCAATCCA CTCTTTTATC AATGACCCAC AGCTGCAGGT 120
GGATTTCTAC ACTGACATGG ATGAGGATTC AGATATTGCC TTCCGTTTCC GAGTGCACTT 180
TGGCAATCAT GTGGTCATGA ACAGGCGTGA GTTTGGGATA TGGATGTTGG AGGAGACAAC 240
AGACTACGTG CCCTTTGAGG ATGGCAAACA ATTTGAGCTG TGCATCTACG TACATTACAA 300
TGAGTATGAG ATAAAGGTCA ATGGCATACG CATTTACGGC TTTGTCCATC GAATCCCGCC 360
ATCATTTGTG AAGATGGTGC AAGTGTCGAG AGATATCTCC CTGACCTCAG TGTGTGTCTG 420
CAATTGAGGG AGATGATCAC ACTCCTCATT GT 452

(2)SEQ ID NO:4的信息:

(i)序列特征:

(A)长度: 138个氨基酸

(B)类型: 氨基酸

(D)拓扑结构: 线性

(ii)分子类型: 多肽

(xi)序列描述: SEQ ID NO:4

Met Ser Leu Asn Arg Ala Tyr Lys Leu Pro Val Ser Leu Ser Val 15
Gly Ser Cys Val Ile Ile Lys Gly Thr Pro Ile His Ser Phe Ile 30
Asn Asp Pro Gln Leu Gln Val Asp Phe Tyr Thr Asp Met Asp Glu 45
Asp Ser Asp Ile Ala Phe Arg Phe Arg Val His Phe Gly Asn His 60
Val Val Met Asn Arg Arg Glu Phe Gly Ile Trp Met Leu Glu Glu 75
Thr Thr Asp Tyr Val Pro Phe Glu Asp Gly Lys Gln Phe Glu Leu 90
Cys Ile Tyr Val His Tyr Asn Glu Tyr Glu Ile Lys Val Asn Gly 105
Ile Arg Ile Tyr Gly Phe Val His Arg Ile Pro Pro Ser Phe Val 120
Lys Met Val Gln Val Ser Arg Asp Ile Ser Leu Thr Ser Val Cys 135
Val Cys Asn 138

(2)SEQ ID NO:5的信息

(i)序列特征

(A)长度: 29碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:5:

CAACGTCGAC ATGTCTCTTA ACCGTGCAT 29

(2)SEQ ID NO:6的信息

(i)序列特征

(A)长度: 29碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:6:

TAGGAAGCTT TCAATTGCAG ACACACACT 29

(2)SEQ ID NO:7的信息

(i)序列特征

(A)长度: 29碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:7

CAACAAGCTT ATGTCTCTTA ACCGTGCAT 29

(2)SEQ ID NO:8的信息

(i)序列特征

(A)长度: 29碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:8

TAGGGGATCC TCAATTGCAG ACACACACT 29

(2)SEQ ID NO:9的信息

(i)序列特征

(A)长度: 23碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:9:

GCCTGGGTGA CAGAGCAAGA ACC 23

(2)SEQ ID NO:10的信息

(i)序列特征

(A)长度: 23碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:10

CTGGGAATCC CATGGTCAGG TAG 23

(2)SEQ ID NO:11的信息:

(i)序列特征:

(A)长度: 662bp

(B)类型: 核酸

(C)链性: 双链

(D)拓扑结构: 线性

(ii)分子类型: cDNA

(xi)序列描述: SEQ ID NO:11

GCCTGGGTGA CAGAGCAAGA ACCTATCTCA AAAGTACAGA AAAATCATCT CATCTACTTG 60
TAGTCATCAT AGAAATCAAT CATTCCCTCC AGTTATGTCC CTGACCCACA AGCTTCATTT 120
GTGCAAGTAC TGGGGCTGTG CTCTCAGTAG TGTGTGCCCC TTCTTGGAAG GATGTCCATG 180
GCCCTTGATG ATAGTGCCAT ACAAACTGCC TGTGTCTTTG TCTGTTGGTT CTTGCGTGAT 240
AATCAAAGGG ACACCAATCC ACTCTTTTAT CAATGACCCA CAGCTGCAGG TGGATTTCTA 300
CACTGACATG GATGAGGATT CAGATATTGC CTTCCGTTTC CGAGTGCACT TTGGCAATCA 360
TGTGGTCATG AACAGGCGTG AGTTTGGGAT ATGGATGTTG GAGGAGACAA CAGACTACGT 420
GCCCTTTGAG GATGGCAAAC AATTTGAGCT GTGCATCTAC GTACATTACA ATGAGTATGA 480
GATAAAGGTC AATGGCATAC GCATTTACGG CTTTGTCCAT CGAATCCCGC CATCATTTGT 540
GAAGATGGTG CAAGTGTCGA GAGATATCTC CCTGACCTCA GTGTGTGTCT GCAATTGAGG 600
GAGATGATCA CACTCCTCAT TGTTGAGGAA TCCCTCTTTC TACCTGACCA TGGGATTCCC 660
AG 662

(2)SEQ ID NO:12的信息:

(i)序列特征:

(A)长度: 167个氨基酸

(B)类型: 氨基酸

(D)拓扑结构: 线性

(ii)分子类型: 多肽

(xi)序列描述: SEQ ID NO:12

Met Ser Leu Thr His Lys Leu His Leu Cys Lys Tyr Trp Gly Cys 15
Ala Leu Ser Ser Val Cys Pro Phe Leu Glu Gly Cys Pro Trp Pro 30
Leu Met Ile Val Pro Tyr Lys Leu Pro Val Ser Leu Ser Val Gly 45
Ser Cys Val Ile Ile Lys Gly Thr Pro Ile His Ser Phe Ile Asn 60
Asp Pro Gln Leu Gln Val Asp Phe Tyr Thr Asp Met Asp Glu Asp 75
Ser Asp Ile Ala Phe Arg Phe Arg Val His Phe Gly Asn His Val 90
Val Met Asn Arg Arg Glu Phe Gly Ile Trp Met Leu Glu Glu Thr 105
Thr Asp Tyr Val Pro Phe Glu Asp Gly Lys Gln Phe Glu Leu Cys 120
Ile Tyr Val His Tyr Asn Glu Tyr Glu Ile Lys Val Asn Gly Ile 135

Arg Ile Tyr Gly Phe Val His Arg Ile Pro Pro Ser Phe Val Lys 150
Met Val Gln Val Ser Arg Asp Ile Ser Leu Thr Ser Val Cys Val 165
Cys Asn 167

(2)SEQ ID NO:13的信息

(i)序列特征

(A)长度: 29碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:13:

CAACGGATCC ATGTCCCTGA CCCACAAGC 29

(2)SEQ ID NO:14的信息

(i)序列特征

(A)长度: 29碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:14:

TAGGGTCGAC TCAATTGCAG ACACACACT 29

(2)SEQ ID NO:15的信息

(i)序列特征

(A)长度: 29碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:15

CAACGGATCC ATGTCCCTGA CCCACAAGC 29

(2)SEQ ID NO:16的信息

(i)序列特征

(A)长度: 29碱基

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO:16

TAGGGAATTC TCAATTGCAG ACACACACT 29

说 明 书 附 图

```
CLC4A   MSLNR----------------------------AYKLPVSLSVGSCVII   21
CLC4B   MSLTHKLHLCKYWGCALSSVCPFLEGCPWPLMIVPYKLPVSLSVGSCVII   50
hCLC    MSL--------------------------LPVPYTEAASLSTGSTVTI   22
mCLC    ------------------------------------------------   0
pCLC    ------------------------------------------------   0

CLC4A   KGTPIHSFINDPQLQVDFYTDMDEDSDIAFRFRVHFGNHVVMNRREFGIW   71
CLC4B   KGTPIHSFINDPQLQVDFYTDMDEDSDIAFRFRVHFGNHVVMNRREFGIW   100
hCLC    KGRPLVCFLNEPYLQVDFHTEMKEESDIVFHFQVCFGRRVVMNSREYGAW   72
mCLC    ----------EPYLQVDFHTEMKEDSDIAFHSRVYFGHWVVMNSRVNGAW   40
pCLC    ----------EPYLQVDFHTEMKEESGIAFHFQVHFGCYVVMNSREYGAW   40
                 .* *****.*.*.*.*.*.*. .* **  ****.*  * *

CLC4A   MLEETTDYVPFEDGKQFELCIYVHYNEYEIKVNGIRIYGFVHRIPPSFVK   121
CLC4B   MLEETTDYVPFEDGKQFELCIYVHYNEYEIKVNGIRIYGFVHRIPPSFVK   150
hCLC    KQQVESKNMPFQDGQEFELSISVLPDKYQVMVNGQSSYTFDHRIKPEAVK   122
mCLC    QYEVTCHNMPFQDGKPFNLSISVPPDKY----------------------   68
pCLC    KKPVESKNMPFQDGQEFDLSISVLPDKY----------------------   68
              . . .**.**. *.*.* *  ..*

CLC4A   MVQVSRDISLTSVCVC---N   138
CLC4B   MVQVSRDISLTSVCVC---N   167
hCLC    MVQVWRDISLTKFNVSYLKR   142
mCLC    --------------------   68
pCLC    --------------------   68
```

共有长度: 170

[1] CLC4A: 138 个氨基酸
[2] CLC4B: 167 个氨基酸
[3] hCLC: 142 个氨基酸
[4] mCLC: 68 个氨基酸
[5] pCLC: 68 个氨基酸

图 1

Human charcot-l yd n crystal 4, its coding s quence, its preparing proc ss and its application

Patent Number: CN1302876
Publication date: 2001-07-11
Inventor(s): YU LONG (CN); ZHANG HONGLAI (CN); ZHAO YONG (CN)
Applicant(s): UNIV FUDAN (CN)
Requested Patent: CN1302876
Application Number: CN19990119883 19991028
Priority Number(s): CN19990119883 19991028
IPC Classification: C12N15/12; C12N15/63; C07K14/435; A61K39/395; C12Q1/68
EC Classification:
Equivalents:

Abstract

The present invention discloses the cDNA sequence of human Charcot-Leyden crystal 4 (CLC4). The protein coded by said cDNA is the homolog of said CLC protein. The polypeptide coded by said polynucleotide sequence, the application of said polynucleotide and polypeptide, and the process for preparing said polynucleotide and polypeptide are also disclosed.

Data supplied from the **esp@cenet** database - I2